From Noise Modeling to Blind Image Denoising

Fengyuan Zhu[1], Guangyong Chen[1], and Pheng Ann Heng[1,2]

[1] Department of Computer Science and Engineering, The Chinese University of Hong Kong
[2]Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences

Abstract

Traditional image denoising algorithms always assume the noise to be homogeneous white Gaussian distributed. However, the noise on real images can be much more complex empirically. This paper addresses this problem and proposes a novel blind image denoising algorithm which can cope with real-world noisy images even when the noise model is not provided. It is realized by modeling image noise with mixture of Gaussian distribution (MoG) which can approximate large varieties of continuous distributions. As the number of components for MoG is unknown practically, this work adopts Bayesian nonparametric technique and proposes a novel Low-rank MoG filter (LR-MoG) to recover clean signals (patches) from noisy ones contaminated by MoG noise. Based on LR-MoG, a novel blind image denoising approach is developed. To test the proposed method, this study conducts extensive experiments on synthesis and real images. Our method achieves the state-of-the-art performance consistently.

1. Introduction

Image denoising is an important problem in the area of computer vision and image processing. It is not only a useful low-level image processing tool to provide high-quality image, but also an important pre-processing step for many high-level visual problems including digital entertainment, object recognition, image segmentation and remote sensing imaging. It aims to recover the clean image $\hat{X}$ from its noisy observation X which is contaminated by noise E with

$$X = \hat{X} + E. \tag{1}$$

Estimating $\hat{X}$ from X is an inverse problem. Many algorithms [11, 31, 36, 29, 30, 4, 8, 14, 9, 34, 18, 38] were proposed in recent years to solve this problem with good performance. Most of these approaches assume that the noise E follows homogeneous white Gaussian distribution. This assumption seems reasonable as some other noise can be converted into Gaussian noise (e.g., Poisson noise can be converted to white Gaussian noise via Anscombe transform [1]). However, in real camera systems, the noise has various sources (e.g., dark current noise, short noise, thermal noise, quantization noise) and can be much more complex. For example, Tsin [32], Liu [17] and Lebrun [15] stated that the noise model of empirical noisy images captured by CCD camera system can be spatial, signal and frequency dependent, and even non-Gaussian empirically. As a result, the noise model in real images can be much more complex than the one-parameter homogeneous white Gaussian distribution. To cope with noisy images contaminated by noise with various unknown noise models, it is important to develop a robust "blind image denoising" algorithm which can adaptively estimate the noise model from a noisy observation, as well as recover the clean image.

However, to our best knowledge, only a few approaches were proposed to address the blind image denoising. Portilla [24, 25] generated the classical BLS-GSM denoising method [26] which models patches at each scale in the wavelet domain with a Gaussian scale mixture (GSM). Instead of assuming that the noise is homogeneous white Gaussian distributed, this method estimates the noise on each wavelet subband with a zero-mean correlated Gaussian model. Portilla proposed a method to estimate the covariance matrix of noise and applied Bayesian least square to recover the clean image. Liu [17] proposed a segmentation-based blind image denoising algorithm for JPEG images. This method assumes the noise to be intensity dependent Gaussian distributed. Taking the advantage that natural image is piecewise-constant, this approach first segments an input noisy image into small piecewise-constant areas, then introduces a so-called "noise level function" (NLF) to model the relation between the intensity of a pixel and the noise level associated with it. After the noise model is estimated, the method recovers the clean image with a conditional random field. Recently, Lebrun [15] introduced a new approach called "multiscale image blind de-

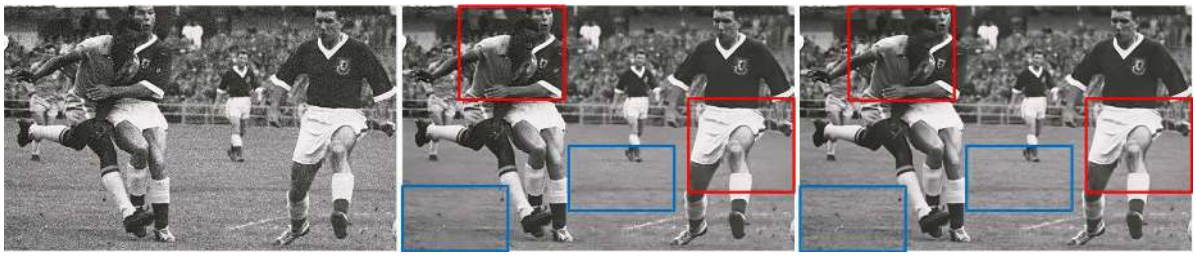

Figure 1. From left to right: an old photo "Pele" with noise; the denoising result of [15] and ours. Zoom in for better visualization.

noising". This algorithm is an adaption of the Non-local Bayes approach (NL-Bayes) [14] which achieves good performance on image denoising with known homogeneous white Gaussian noise. Different from the NL-Bayes algorithm, the noise model of each patch and its nearby patches is assumed to be zero-mean correlated Gaussian distributed instead of white Gaussian with known intensity. This extension is quite similar to Portilla's one [24, 25]. The noise covariance matrix of the patches around each given patch is estimated by [7] which is an intensity-frequency dependent noise estimation algorithm for JPEG images.

All of the above mentioned algorithms are the conjunction of a thorough method to estimate the noise model followed by an adapted denoising method. Specifically, these approaches all first divide patches (or pixels) into groups. They further assume that the noise on each group is homogeneous Gaussian distributed. After that, they propose different noise estimation method to obtain the Gaussian parameters for each group and recover the clean image with adaptive denoising methods. The noise models of these approaches are much more general than the homogeneous white Gaussian distribution. However, when the camera information and the image capture environments are unknown, the dependency relation between the image and noise is not explicit and can be very complex. And the noise can be heterogeneous and even non-Gaussian for empirical noisy images. It can be very difficult to propose an optimal patch (or pixel) grouping scheme to ensure the noise on each group to be homogeneous Gaussian distributed. Thus, conventional noise models can still fail to handle empirical noise. Fig. 1 demonstrates the performance of [15] on an old photo "Pele". Due to the lack of flexibility of the noise model, [15] can only eliminate some but not complete noise in some areas (e.g., within the red boxes) and over-smooth details in some other areas (e.g., within the blue boxes).

To tackle the above mentioned problem, this paper proposes a novel non-local blind image denoising algorithm. Similar to conventional approaches, the method also divides the patches into groups. Differently, our approach relaxes the assumption that the noise of certain group of patches are homogeneous Gaussian distributed by proposing a model that is more general for empirical noise on noisy images. As the noise on an image can be heterogeneous and even non-Gaussian distributed, the noise model should be flexible enough to cover large varieties of distributions. Thus, this work introduces mixture of Gaussian (MoG) [20] to model the noise for each group of patches. Since MoG can not only approximate any continuous distribution efficiently [2], but also model multi-modal data, it is more general and capable of adapting the empirical noise on real-world noisy images. Similar noise modeling strategy can also be found in [37, 21]. Taking the advantage that natural signals (e.g., clean image patches) are low-rank representable [6, 12, 35, 33], we assume the clean patches within each group are low-rank representable. To recover the clean patches from noisy one, this study further proposes **L**ow **R**ank **MoG** filter (LR-MoG filter) to decompose MoG noise from low-rank representable signals. As the complexity of each MoG noise (number of components for MoG) is usually unknown, this work introduces Bayesian nonparametric techniques to build LR-MoG filter. This work treats LR-MoG filter in fully-Bayesian way and infers both the complexity of noise and the clean signals with variational Bayesian method [3]. Fig. 1 shows the performance of our algorithm on "Pele" compared with the state-of-the-art blind denoising approach [15]. Our algorithm can eliminate the noise efficiently and preserve the detailed features. We further conduct extensive experiments to test our method on more real and synthesized noisy images with different noise model. Compared with other state-of-the-art approaches, our method achieves superior performance consistently.

2. Background

This section introduces the background information of Dirichlet process and its construction, which will be used to construct the LR-MoG filter.

2.1. Dirichlet Process

The Dirichlet process (DP) is a distribution over distributions, *i.e.*, each draw from a DP is itself a distribution. The DP is parameterized by a base distribution H and a concentration parameter α. A nice feature favored by DP is that a drawn G from a DP is discrete with probability one [10] and the dimension of G is potentially infinite. Since DP generates discrete distributions on continuous parameter space, it

is usually used as prior to construct nonparametric mixture models which is represented as the convex combination of component distributions and the number of components can be decided by data.

2.2. Stick-breaking Construction

As H is diffuse, the representations of G at the granularity DP must be represented via construction methods, which is necessary for inference.

The stick-breaking construction [28] is one of these methods to explicitly represent draws from DP. With a stick-breaking construction, one can directly work with G before drawing θ. Sethuraman proved that a draw G from $DP(\alpha, H)$ can be described as

$$\begin{array}{cc} v_i \sim \text{Beta}(1,\alpha), & \pi_i \sim v_i \prod_{j=1}^{i-1}, \\ \theta_i \sim H, & G = \sum_{i=1}^{\infty} \pi_i \delta_{\theta_i}. \end{array} \tag{2}$$

Here, $\text{Beta}(a, b)$ is a Beta distribution with parameters a and b. And δ_{θ_i} is the Dirac probability measure concentrated at θ_i. $\boldsymbol{\pi}$ are the stick lengths, and it is almost sure that $\sum_{i=1}^{\infty} \pi_i = 1$. The stick-breaking construction indicates the discreteness of G as well.

3. LR-MoG Filter

This section introduces the LR-MoG filter to recover low-rank signals from noisy ones contaminated by MoG noise. Mathematically, let $X = [x_1 \cdots x_n]$ be a $d \times n$ matrix and $x_i \in \mathcal{R}^d$, $i \in \{1, \cdots, n\}$, be a noisy signal. And X can be decomposed as

$$X = \hat{X} + E. \tag{3}$$

Here, $\hat{X} = [\hat{x}_1 \cdots \hat{x}_n]$, where $\hat{x}_i$ is the underlying clean signal of the noisy one $\hat{x}_i$ and $\hat{X}$ is low-rank representable. $E = [e_1 \cdots e_n]$, where e_i is the noise on x_i and follows a MoG distribution. LR-MoG is proposed to recover $\hat{X}$ from observed X. To decompose the signal from noise, we introduce Bayesian approach by introducing priors on $\hat{X}$ and E. Then the signal is recovered via variational inference.

3.1. Modeling Low-rank Component

We first model the low-rank component $\hat{X}$. We decompose each vector $\hat{x}_i$, where $i \in \{1, \cdots, n\}$, with $\hat{x}_i = A\hat{y}_i + u$. A is a $d \times d$ matrix and u is the mean vector of $\{\hat{x}_i\}_{i=1}^n$. Similar decomposition methods can also be found in the area of dimensional reduction [2]. Thus, the matrix $\hat{X}$ can be decomposed with

$$\hat{X} = A\hat{Y}^T + U, \tag{4}$$

where $\hat{Y} = [\hat{y}_1 \cdots \hat{y}_n]^T$, $U = [u \cdots u]$ and the column vectors in matrix $A\hat{Y}$ are zero mean. Let $\text{rank}(M)$ be the rank of a matrix M. We introduce the Gaussian prior on u with

$$u \sim \mathcal{N}(\cdot|u_0, K_0). \tag{5}$$

Here, $\mathcal{N}(\cdot|\mu, \Sigma)$ is a normal distribution with mean μ and covariance Σ. To model the low-rank property of $\hat{X}$, it is intuitive to let matrix $A\hat{Y}^T$ be low-rank representable, because $\text{rank}(U) = 1$ and $\text{rank}(\hat{X}) = \text{rank}(A\hat{Y}^T + U) \leq \text{rank}(A\hat{Y}) + 1$.

To model the low-rank property of $A\hat{Y}^T$, a simple way is to introduce the trace-norm prior [6, 5] on it which can be well approximated by the ARD [27]. Here, we adopt the soft trace-norm prior on $A\hat{Y}^T$ with

$$p(A) \propto \exp(-\frac{1}{2}\text{tr}(AC_A^{-1}A^T)) \tag{6}$$

$$p(\hat{Y}) \propto \exp(-\frac{1}{2}\text{tr}(\hat{Y}C_{\hat{Y}}^{-1}\hat{Y}^T)), \tag{7}$$

where $\text{tr}(\cdot)$ represents the trace of a matrix. C_A and $C_{\hat{Y}}$ are set to be diagonal positive semidefinite with

$$C_A = \text{diag}_d\{c_{a_1}^2, \cdots, c_{a_d}^2\} \tag{8}$$

$$C_{\hat{Y}} = \text{diag}_d\{c_{y_1}^2, \cdots, c_{y_d}^2\}. \tag{9}$$

Here, $\text{diag}_d\{c_1^2, \cdots, c_n^2\}$ represents a $d \times d$ diagonal matrix with diagonal items $c_1^2, \cdots, c_n^2$. Let $\hat{a}_j$ be the jth row vector of matrix A, we can derive the specific formula of Eq. (6) with

$$p(A) \propto \exp(-\frac{1}{2}\text{tr}(AC_A^{-1}A^T)) \tag{10}$$

$$\propto \prod_{j=1}^{d} \exp(-\frac{1}{2c_{a_j}^2}\hat{a}_j^T\hat{a}_j) \tag{11}$$

$$= \prod_{j=1}^{d} \mathcal{N}(\hat{a}_j|0, c_{a_j}^2 I). \tag{12}$$

Eq. (12) is derived via Gaussian integral. Similarly, Eq. (7) can be derived with $p(\hat{Y}) = \prod_{j=1}^{d} \mathcal{N}(\hat{y}_j|0, c_{y_j}^2 I)$.

3.2. Modeling MoG Noise

To model the complex noise on practical noisy images, we propose MoG distribution for noise modeling. Since the number of components is not provided, we use Bayesian nonparametric technique and introduce the Dirichlet process prior to the MoG. Each Gaussian component is represented by $\mathcal{N}(\cdot|\mu_i, \Sigma_i)$ with mean vector μ_i and covariance matrix Σ_i.

For the base distribution H, we introduce conjugate prior over μ_i and Σ_i with

$$\mu_i \sim \mathcal{N}(\cdot|\mu_0, \Omega_0) \tag{13}$$

$$\Sigma_i \sim \text{iWishart}(\cdot|a_0, B_0). \tag{14}$$

Here, iWishart$(\cdot|a_0, B_0)$ is the inverse-Wishart distribution with a_0 as the degree of freedom and B_0 as scale matrix. The DP mixture is built via stick-breaking construction in Eq. (2) and generative process of e_i is as follows

1. Draw $v_t \sim$ Beta$(1, \alpha)$, for $t = \{1, 2, \cdots\}$ and $\pi_t = v_t \prod_{j=1}^{t-1}(1 - v_j)$
2. Draw $\mu_t \sim \mathcal{N}(\cdot|\mu_0, \Omega_0)$ and $\Sigma_t \sim$ iWishart$(\cdot|a_0, B_0)$
3. For each e_i:
   (a) Draw $z_i \sim$ Mult$(\boldsymbol{\pi})$
   (b) Draw $e_i \sim \mathcal{N}(\cdot|\mu_{z_i}, \Sigma_{z_i})$.

Following the setting in previous works [17, 15, 4], we set the noise to be zero mean. Note that, we only constrain that the marginal mean of all Gaussian components for the MoG noise is zero and the mean of each component is unconstrained. This noise model can well approximate all continuous distributions even when they are heterogeneous and non-Gaussian distributed. Thus, the proposed noise model is still much more general than conventional ones [15, 4, 17, 26, 24].

3.3. Remarks on LR-MoG

There are other algorithms which were proposed to decompose low-rank signals from MoG noise including [21] and [37]. However, our approach is different in three folds. Firstly, our method is the only approach to adopt Bayesian nonparametric technique with DP prior to learn the number of components for MoG automatically. In contrast, [21] tunes the component number manually and [37] uses Dirichlet distribution prior and choose the component number with a heuristic scheme. Secondly, [21] and [37] all assume each noise component to be white Gaussian distributed which is relaxed in LR-MoG. Thirdly, the low-rank component modeling for these approaches are different as well.

3.4. Variational Inference and Signal Recovery

Let Θ be all the parameters in the above model including A, $\hat{Y}$ and parameters of Gaussian components for noise; and $\Delta = \{u_0, K_0, C_A, C_{\hat{Y}}, \mu_0, \Omega_0, a_0, B_0\}$ be the hyper-parameters of the priors on Θ. With the observed matrix X, the posterior distribution of Θ given X and Δ is

$$p(\Theta|X, \Delta) = \frac{p(X|\Theta, \Delta)p(\Theta|\Delta)}{p(X|\Delta)}. \tag{15}$$

Empirically, the analytical form of $p(\Theta|X, \Delta)$ is very complex and can be even computationally intractable. Thus, we propose Variational Bayesian (VB) inference technique to obtain a tight approximation of the posterior. Mathematically, the VB seeks a distribution $q(\Theta)$ to minimize the variational function as follows,

$$\mathcal{E}^{VB} = \int q(\Theta) \ln \frac{q(\Theta)}{p(X|\theta)p(\theta|\Delta)} \mathrm{d}\Theta \tag{16}$$

$$= \mathrm{KL}(q(\Theta)||p(\Theta|X, \Delta)) - \ln p(X|\Delta). \tag{17}$$

Here, the first term of Eq. (17) is the KL divergence between $q(\Theta)$ and $p(\Theta|X, \Delta)$; the second term is a constant with respect to Θ.

With the model built in Sec. 3.1 and 3.2, the prior on Θ is given as:

$$p(\Theta|\Delta) = p(Z|V)p(V|\alpha)p(A|C_A)p(\hat{Y}|C_{\hat{Y}}) \\ p(u|u_0, K_0) \prod_{t=1}^{T} p(\mu_t|\mu_0, \Omega_0)p(\Sigma_t|a_0, B_0). \tag{18}$$

Here, $Z = \{z_1, \cdots, z_n\}$, $V = \{v_1, \cdots, v_T\}$, where T is the number of components for MoG noise, which is inferred with truncation of DP [13]. And the likelihood $p(X|\Theta, \Delta) = \prod_{i=1}^{n} p(x_i|\Theta, \Delta)$. Here, we adopt the mean field VB [3] to factorize $q(\Theta)$ as

$$q(\Theta) = q(Z)q(V)q(u)q(A)q(\hat{Y}) \prod_{t=1}^{T} q(\mu_t)q(\Upsilon_t), \tag{19}$$

where $q(Z) = \prod_{i=1}^{n} q(z_i)$ and $q(V) = \prod_{t=1}^{T} q(v_t)$. To simplify the notations, $q(z_i = t)$ is denoted as $q_i(t)$ and $\langle g(x) \rangle$ denotes the expectation of $g(x)$ over x. Then the posterior of all parameters in Θ can be approximated with VB by iteratively updating parameters with Sec. 3.4.2 and Sec. 3.4.1.

3.4.1 MoG Noise Inference

To infer the noise component in Eq. (3), we need to approximate the posterior of $\{Z, V\} \cup \{\mu_t, \Sigma_t\}_{t=1}^{T}$ with the prior distributions in Eq. (18).

Let $\xi_t = \sum_{i=1}^{n} q_i(t)$. For each $v_t \in V$, the posterior distribution is still Beta distributed with Beta(α_t, β_t), where

$$\alpha_t = \xi_t + 1, \quad \beta_t = \alpha + \textstyle\sum_{j=t+1}^{T} \xi_j. \tag{20}$$

For updating the posterior of Z, we have

$$q_i(t) = \frac{\rho_i(t)}{\sum_{j=1}^{T} \rho_i(j)}, \tag{21}$$

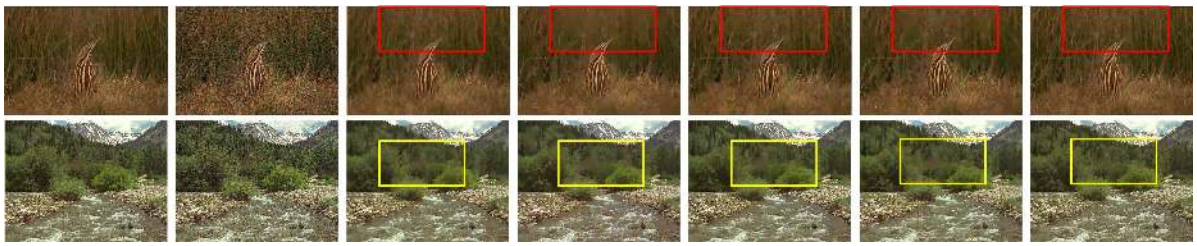

Figure 2. Denoising results of '43033' from BSDS 500 and 'I13' from TID 500. From left to right: clean images, noisy observations with $\sigma = 40$, results of K-SVD, SURE-GMM, BM3D, NL-Bayes and ours. Our approach can better preserve details marked by the boxes. Zoom in for better visualization.

where $\rho_i(j) = \exp(\tau^i_{j,1} + \tau^i_{j,2})$ with

$$\begin{aligned}\tau^i_{j,1} &= \psi(\alpha_j) + \psi(\beta_j)\\ \tau^i_{j,2} &= -\frac{1}{2}\mathrm{tr}\{\langle\Sigma_j^{-1}\rangle[\langle uu^T\rangle + 2\langle\mu_j\rangle\langle u^T\rangle + \langle\mu_j\mu_j^T\rangle + x_i x_i^T\\ &\quad + \langle y_i y_i^T\rangle\langle A^T A\rangle - 2\langle A\rangle\langle y_i\rangle(x_i - \langle u\rangle - \langle\mu_j\rangle)^T\\ &\quad - 2x_i(\langle u\rangle + \langle\mu_j\rangle)^T] + \langle\ln|\Sigma_t|\rangle\}.\end{aligned} \tag{22}$$

Here, $\psi(\cdot)$ denotes the digamma function.

With the Eq. (21), we can infer the posterior of μ_t and Σ_t for $t \in \{1, \cdots, T\}$. For μ_t, the posterior $q(\mu_t)$ is still Gaussian distributed with:

$$\begin{aligned}\langle\mu_t\rangle &= [\Sigma_0^{-1} + \sum_{i=1}^{n} q_i(t)\langle\Sigma_t^{-1}\rangle]^{-1}[\Sigma_0^{-1}\mu_0\\ &\quad + \sum_{i=1}^{n} q_i(t)\langle\Sigma_t^{-1}\rangle(x_i - \langle u\rangle - \langle A\rangle\langle y_i\rangle)]\\ \langle\mu\mu_t^T\rangle &= [\Sigma_0^{-1} + \sum_{i=1}^{n} q_i(t)\langle\Sigma_t^{-1}\rangle]^{-1} + \langle\mu_t\rangle\langle\mu_t\rangle^T.\end{aligned} \tag{23}$$

And the posterior of Σ_t is still inverse-Wishart distributed as iWishart$(\cdot|a_t, B_t)$ with parameters

$$\begin{aligned}a_t &= a_0 + \xi_t\\ B_t &= B_0 + \frac{1}{2}\sum_{i=1}^{n} q_i(t)(\langle uu^T\rangle + 2\langle\mu_t\rangle\langle u\rangle^T + \langle\mu_t\mu_t^T\rangle\\ &\quad + x_i x_i^T + \langle y_i y_i^T\rangle\langle A^T A\rangle - 2(x_i - u - \mu_t)^T\langle A\rangle\langle y_i\rangle\\ &\quad - 2(\langle u\rangle + \langle\mu_t\rangle)^T x_i).\end{aligned} \tag{24}$$

With the properties of inverse-Wishart distribution, we have

$$\begin{aligned}\langle\Sigma_t^{-1}\rangle &= a_t(B_t)^{-1}\\ \langle\ln|\Sigma_t|\rangle &= \frac{1}{\psi(a_t/2) + d\ln 2 + \ln|B_t^{-1}|}.\end{aligned} \tag{25}$$

3.4.2 Low-rank Component Inference

To infer the low-rank component in Eq. (3), we need to estimate the posterior of u, A and $\hat{Y}$ with the prior distributions in Eq. (18).

We first solve the posterior of the mean vector u. Since the noise is assumed to be zero mean marginally, u corresponds to the mean vector of the observed signals. The posterior of u is still Gaussian distributed with

$$\begin{aligned}\langle u\rangle &= [K_0^{-1} + \sum_{i=1}^{N}\sum_{t=1}^{T} q_i(t)\langle\Sigma_t^{-1}\rangle]^{-1}[K_0^{-1}u_0\\ &\quad + \sum_{i=1}^{N}\sum_{t=1}^{T} q_i(t)\langle\Sigma_t^{-1}\rangle(x_i - \langle\mu_t\rangle - \langle A\rangle\langle y_i\rangle)]\\ \langle uu^T\rangle &= [K_0^{-1} + \sum_{i=1}^{N}\sum_{t=1}^{T} q_i(t)\langle\Sigma_t^{-1}\rangle]^{-1} + \langle u\rangle\langle u\rangle^T.\end{aligned} \tag{26}$$

We turn to infer the posterior of A and write $A = [\hat{a}_1 \cdots \hat{a}_d]$, where each $\hat{a}_j \in \mathcal{R}^d, j \in \{1, \cdots, n\}$ can be considered as the jth basis of A. Let $y_{i,j}$ be the jth component of vector y_i, the posterior distribution over A_t satisfies:

$$\begin{aligned}\langle A\rangle &= [\langle\hat{a}_1\rangle, \ldots, \langle\hat{a}_d\rangle],\\ \langle\hat{a}_j\rangle &= [\frac{1}{c^2_{a_j}}\mathbf{I} + \sum_{i=1}^{n}\sum_{t=1}^{T} q_i(t)\langle\Sigma_t^{-1}\rangle\langle y_{i,j}\rangle^2]^{-1}[\sum_{i=1}^{n}\sum_{t=1}^{T}\\ &\quad q_i(t)\langle\Sigma_t^{-1}\rangle\langle y_{i,j}\rangle(x_i - \langle\mu_t\rangle - \langle u\rangle - \sum_{v\neq j}^{d}\langle y_{i,v}\rangle\langle\hat{a}_v\rangle)]\\ \langle\hat{a}_j^T\hat{a}_j\rangle &= [\frac{1}{c^2_{a_j}}\mathbf{I} + \sum_{i=1}^{n}\sum_{t=1}^{T} q_i(t)\langle\Sigma_t^{-1}\rangle\langle y_{i,j}\rangle^2]^{-1} + \langle\hat{a}_j\rangle^T\langle\hat{a}_j\rangle.\end{aligned} \tag{27}$$

And $\langle A^T A\rangle$ is a $d \times d$ matrix whose element $\langle A^T A\rangle_{i,j}$ is $\langle\hat{a}_i^T\hat{a}_j\rangle$ when $i = j$, and $\langle\hat{a}_i\rangle^T\langle\hat{a}_j\rangle$ when $i \neq j$. The inference of $\hat{Y}$ is very similar with that of A, which is omitted in this paper.

3.5. Signal Recovery

With noisy observations X, the parameters of LR-MoG filter can be efficiently estimated with the variational inference in Sec. 3.4. And a clean signal $\hat{x}_i$ can be recovered with

$$\hat{x}_i = \langle A \rangle \langle y_i \rangle + \langle u \rangle. \tag{28}$$

4. Blind Image Denoising with LR-MoG

Based on the LR-MoG filter discussed in Sec. 3, we propose a novel scheme for blind image denoisnig. Specifically, an input noisy image is divided into overlapping patches and each patch x_i is treated as a noisy signal. For each patch x_i, we search its similar patches non-locally with nearest neighbor search [2]. This step is the same as conventional non-local means algorithms [4, 8, 14].

We combine x_i and its nearby patches to a matrix X as in Eq. (3). Then we use the LR-MoG to recover the clean patches (signals) within X. When all the patches are processed, we aggregate the denoised patches into the clean image with the same scheme as in [8].

5. Experiments

To test the proposed method, we conduct extensive experiments on both synthesized and real noisy images with comparison of state-of-the-art methods. Our algorithm achieves best performance in all the experiments.

5.1. Parameter Setting

We implement the proposed method in Matlab 2014b with the parameters set as follows through this work (without tunning): the patch size is set to be 8×8 and when searching similar patches, we use the ϵ-nearest neighbor scheme with $\epsilon = 4000$. For the hyper-parameters for LR-MoG filtering, we adopt non-informative manner [2] to minimize their influence on the posterior distributions. Specifically, we set $K_0, C_A, C_{\hat{Y}}, \Omega_0, B_0$ to be $\mathbf{I}$, u_0, μ_0 to be zero vector, and a_0 to be 64.

5.2. Synthesis Noisy Image

We first test our method on synthesized noisy images because the ground truth is available for objective evaluations. We use the images from two datasets: TID 2008 [23] and BSDS 500 [19]. We use the code or executable codes released by the authors of competitive methods. The performance is measured by peak signal to noise ratio (PSNR) and the structural similarity (SSIM) [35].

5.2.1 AWGN Noise

We first test our method on noisy images contaminated by homogeneous white Gaussian noise, which has been studied for decades in the problem of non-blind Image denoising. As the white Gaussian noise is also a special case of our noise model, our approach can handle these images efficiently. We compare the proposed method with 4 state-of-the-art methods: K-SVD [9], SURE-guided GMM [34], BM3D [8] and NL-Bayes [14] on images contaminated by noise with intensity from 10 to 100. As these methods are non-blind, we provide the real noise intensity to them while let our method to learn the noise model automatically. Table 1 shows the average quantitative performance of the methods on the two datasets. Our approach has better performance when the noise is high ($\sigma > 30$) and the results are competitive when $\sigma \leq 30$. Fig. 2 shows the denoising results on two noisy images with $\sigma = 40$ of all the methods. It can be observed that our method can well eliminate the noise and better preserve the details (grasses in '43033' and trees in 'I13'). It shows that our method can well cope with images contaminated by AWGN with different intensity.

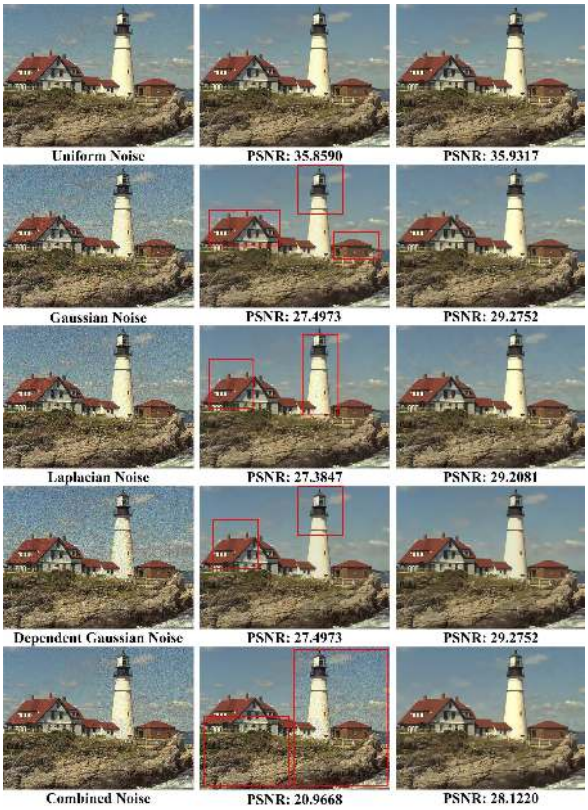

Figure 3. Comparison of proposed method and [15] on 'I21' in TID 2008. From left to write: noisy images (Uniform: $s = 30$, Gaussian: $\sigma = 30$, Laplacian: $\sigma = 30$, Dependent: $b = 4$), results of [15] and our approach. Our method can better eliminate noise while [15] cannot eliminate noise completely as marked by red boxes. Zoom in for better visualization.

Table 1. Performance of different approaches on noisy images with homogeneous white Gaussian noise with different deviation.

	σ	10	20	30	40	50	60	70	80	90	100
Dataset	Method	PSNR									
TID2008	K-SVD	34.74	30.87	28.97	27.89	26.04	24.92	24.45	23.80	23.33	22.84
	SURE-GMM	34.79	31.22	29.23	27.91	26.53	24.89	24.78	23.81	23.70	23.04
	BM3D	**35.17**	31.46	29.28	28.02	26.66	25.39	25.14	24.66	24.11	23.26
	NL-Bayes	35.06	**31.51**	**29.31**	28.04	26.67	**25.43**	25.11	24.51	23.84	23.30
	Ours	35.09	31.39	29.24	**28.06**	**26.71**	25.41	**25.22**	**24.69**	**24.21**	**23.31**
BSDS500	K-SVD	34.24	30.69	28.66	27.69	25.93	25.10	24.22	23.46	23.12	22.67
	SURE-GMM	34.22	31.01	29.02	27.68	26.23	25.22	24.56	23.58	23.53	22.97
	BM3D	**34.72**	**31.19**	29.11	27.51	26.56	**25.68**	24.92	24.39	23.97	23.02
	NL-Bayes	34.69	31.09	29.09	27.53	26.48	25.66	24.79	24.41	23.68	23.07
	Ours	34.68	31.13	**29.10**	**27.81**	**26.63**	25.61	**24.95**	**24.44**	**23.97**	**23.11**
		SSIM									
TID2008	K-SVD	0.947	0.930	0.900	0.861	0.841	0.817	0.798	0.775	0.764	0.743
	SURE-GMM	0.944	0.935	0.904	0.868	0.848	0.818	0.807	0.787	0.770	0.747
	BM3D	**0.968**	**0.938**	0.912	0.874	0.856	0.828	0.816	0.800	0.781	0.764
	NL-Bayes	0.962	0.934	**0.916**	0.875	0.860	0.827	**0.818**	0.794	0.779	0.761
	Ours	0.954	0.931	0.913	**0.876**	**0.860**	**0.830**	0.814	**0.803**	**0.784**	**0.768**
BSDS500	K-SVD	0.938	0.925	0.897	0.856	0.830	0.822	0.790	0.766	0.762	0.741
	SURE-GMM	0.939	0.923	0.899	0.859	0.833	0.827	0.793	0.774	0.768	0.746
	BM3D	**0.961**	**0.931**	0.907	**0.867**	0.848	0.835	0.806	0.792	0.777	0.763
	NL-Bayes	0.953	0.926	0.908	0.866	0.849	0.834	0.810	0.793	0.772	0.764
	Ours	0.954	0.926	**0.912**	0.863	**0.849**	**0.838**	**0.812**	**0.801**	**0.781**	**0.764**

5.2.2 Complex Noise

Practically, the noise model can be much more complex than the AWGN model. Thus, we further test our method on synthesis data with other types of noise. We add the following types of zero-mean noise to clean images to form synthesis noisy images: (1) AWGN noise with $\sigma = 15, 30, 45$; (2) intensity dependent Gaussian noise with $\sigma = \frac{1}{s}x_{i,j}$, where $x_{i,j}$ is the pixel intensities, with $b = 3, 4, 5$; (3) Laplacian noise with $\sigma = 15, 30, 45$; (4) uniform noise of $[-s, s]$ with $s = 15, 30, 45$; (5) combined noise: we divide an image into four parts: left up, right up, left down, right down. Four parts are contaminated by uniform noise with $s = 15$, Gaussian noise with $\sigma = 30$, Laplacian noise with $\sigma = 30$ and intensity dependent Gaussian noise with $b = 4$ respectively, resulting in spatial dependent noise.

We compare our method with the state-of-the-art multi-scale blind image denoising method [15][1] with the code released by the author. Table 2 shows the quantitative results of the two methods. It is clear that our method outperforms [15] by a significant margin. Fig. 3 shows an example of denoising result for comparison. Our method can well eliminate noise and preserve details while [15] cannot eliminate noise efficiently. It shows that our noise model is more flexible to cope with images with different noise model.

[1]We only compare our method with [15] because it is the only blind image denoising approach with released code or executable file. And it achieves the state-of-the-art performance.

Figure 4. From left to right: Two noisy images from [16]; denoising results of [15] and the proposed method; crops of the denoised result for comparison, left of red line: [15], right of red line: proposed method. Zoom in for better visualization.

5.3. Real Noisy Images

We further test our method on real noisy images captured by real camera systems, with comparison of [15]. We first test our method on real images in [22, 16], which were captured by CCD camera. Examples of the denoising results of two methods are shown in Fig. 4. Our method can well eliminate the noise on these images and achieve better results visually. As a comparison, [15] can also eliminate some noise, but not completely. The denoising results for more images can be found in the supplementary files.

We further apply our methods on the application of old photo recovery. The noise on these images is with large grain generally, which is very challenging for denoising. Fig. 5 shows the denoising results of an old photos on s-

Table 2. Performance of the multiscale approach and ours on images contaminated by five different types of noise.

		Gaussian			Heterogeneous			Laplace			Uniform			Comb.
		15	30	45	3	4	5	15	30	45	15	30	45	
Dataset	Method	PSNR												
TID2008	Multiscale	30.27	27.02	25.14	24.07	25.54	26.16	29.78	26.87	24.97	34.22	31.77	28.97	22.12
	Ours	**33.14**	**29.18**	**27.09**	**27.58**	**28.84**	**29.89**	**33.13**	**28.84**	**27.06**	**36.18**	**35.25**	**32.13**	**29.83**
BSDS500	Multiscale	29.47	26.83	25.00	23.87	25.12	25.47	29.63	26.53	24.29	34.22	30.54	28.24	21.82
	Ours	**32.69**	**28.36**	**29.23**	**26.84**	**28.03**	**30.02**	**32.13**	**28.02**	**26.10**	**35.22**	**34.64**	**30.76**	**29.27**
		SSIM												
TID2008	Multiscale	0.913	0.871	0.831	0.782	0.841	0.836	0.918	0.868	0.831	0.952	0.943	0.922	0.802
	Ours	**0.944**	**0.903**	**0.858**	**0.861**	**0.878**	**0.912**	**0.944**	**0.886**	**0.874**	**0.971**	**0.945**	**0.921**	**0.889**
BSDS500	Multiscale	0.912	0.868	0.826	0.778	0.840	0.829	0.912	0.859	0.836	0.952	0.944	0.909	0.802
	Ours	**0.934**	**0.902**	**0.847**	**0.859**	**0.877**	**0.902**	**0.932**	**0.893**	**0.844**	**0.964**	**0.955**	**0.916**	**0.887**

Figure 5. Two noisy old images and crops of denoising results with different methods. Left of red line: [15]; right of red line: proposed method. Zoom in for better visualization.

Figure 6. From left to right: old photos 'young Einstein' and 'Hilbert'; the denoising results of [15] and ours and crops for better visualization (left of the red lines are [15] and right are ours). Zoom in for better visualization.

ports [2]. Our method can better eliminate noise while preserve the details. Fig. 6 and 7 show the denoising results of old photos 'young Einstein', 'Hilbert' and 'Einstein met Tagore' [3]. Still, our method can better eliminate the noise while [15] can only remove some of them. Here, we only show the crops of denoising results in Fig. 5 and 7 for better visualization. It can be observed that our method achieves better visual performance. It shows that our noise model can better cope with noise on real images and our method is a better candidate for real-world denoising problems.

[2]downloaded from `www.nba.com`

[3]downloaded from `https://www.brainpickings.org/`

Figure 7. From left to right: an old photo 'Einstein met Tagore'; the crops of denoising results. Left of the red line are [15] and the right are ours. Zoom in for better visualization.

6. Conclusion

We propose a novel blind image denoising algorithm which can efficiently recover real noisy images whose noise model is complex and unavailable. We assume the clean image patches are low-rank representable and we model the unknown noise with MoG due to its flexibility in approximating different distributions. To recover the clean images and eliminate noise, we develop the LR-MoG filtering which can learn the noise model automatically with Bayesian nonparametric technique and recover the latent low-rank signals efficiently. And the blind denoising scheme is developed based on the LR-MoG. To test our method, we conduct extensive experiments on synthesis and real images. Our method achieves the best performance consistently. It shows that our approach can cope with real-world noisy image recovery tasks efficiently.

Acknowledgments: This work was supported by the National Basic Program of China, 973 Program (Project No. 2015CB351706), the Research Grants Council of Hong Kong (Project No. CUHK412513), and Shenzhen-Hong Kong Innovation Circle Funding Program (Project No. GHP/002/13SZ and SGLH20131010151755080).

References

[1] F. J. Anscombe. The transformation of poisson, binomial and negative-binomial data. *Biometrika*, pages 246–254, 1948.

[2] C. M. Bishop. *Pattern recognition and machine learning*. springer, 2006.

[3] D. M. Blei, M. I. Jordan, et al. Variational inference for dirichlet process mixtures. *Bayesian analysis*, 1(1):121–143, 2006.

[4] A. Buades, B. Coll, and J.-M. Morel. A non-local algorithm for image denoising. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pages 60–65. IEEE, 2005.

[5] E. J. Candès, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *Journal of the ACM (JACM)*, 58(3):11, 2011.

[6] E. J. Candès and B. Recht. Exact matrix completion via convex optimization. *Foundations of Computational mathematics*, 9(6):717–772, 2009.

[7] M. Colom, M. Lebrun, A. Buades, and J. Morel. A non-parametric approach for the estimation of intensity-frequency dependent noise. In *Image Processing (ICIP), 2014 IEEE International Conference on*, pages 4261–4265. IEEE, 2014.

[8] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *Image Processing, IEEE Transactions on*, 16(8):2080–2095, 2007.

[9] M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *Image Processing, IEEE Transactions on*, 15(12):3736–3745, 2006.

[10] T. S. Ferguson. A bayesian analysis of some nonparametric problems. *The annals of statistics*, pages 209–230, 1973.

[11] R. C. Gonzalez and R. E. Woods. Digital image processing. *Prentice Hall*, pages 299–300, 2002.

[12] S. Gu, L. Zhang, W. Zuo, and X. Feng. Weighted nuclear norm minimization with application to image denoising. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 2862–2869. IEEE, 2014.

[13] H. Ishwaran and L. F. James. Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association*, 96(453), 2001.

[14] M. Lebrun, A. Buades, and J.-M. Morel. A nonlocal bayesian image denoising algorithm. *SIAM Journal on Imaging Sciences*, 6(3):1665–1688, 2013.

[15] M. Lebrun, M. Colom, and J.-M. Morel. Multiscale image blind denoising. 2014.

[16] M. Lebrun, M. Colom, and J.-M. Morel. The Noise Clinic: a Blind Image Denoising Algorithm. *Image Processing On Line*, 5:1–54, 2015.

[17] C. Liu, R. Szeliski, S. B. Kang, C. L. Zitnick, and W. T. Freeman. Automatic estimation and removal of noise from a single image. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(2):299–314, 2008.

[18] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman. Non-local sparse models for image restoration. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 2272–2279. IEEE, 2009.

[19] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proc. 8th Int'l Conf. Computer Vision*, volume 2, pages 416–423, July 2001.

[20] G. McLachlan and D. Peel. *Finite mixture models*. John Wiley & Sons, 2004.

[21] D. Meng and F. De la Torre. Robust matrix factorization with unknown noise. In *Computer Vision (ICCV), 2013 IEEE International Conference on*, pages 1337–1344. IEEE, 2013.

[22] G. Petschnigg, R. Szeliski, M. Agrawala, M. Cohen, H. Hoppe, and K. Toyama. Digital photography with flash and no-flash image pairs. *ACM transactions on graphics (TOG)*, 23(3):664–672, 2004.

[23] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti. Tid2008-a database for evaluation of full-reference visual quality assessment metrics. *Advances of Modern Radioelectronics*, 10(4):30–45, 2009.

[24] J. Portilla. Blind non-white noise removal in images using gaussian scale mixtures in the wavelet domain. In *Benelux Signal Processing Symposium*, 2004.

[25] J. Portilla. Full blind denoising through noise covariance estimation using gaussian scale mixtures in the wavelet domain. In *Image Processing, 2004. ICIP'04. 2004 International Conference on*, volume 2, pages 1217–1220. IEEE, 2004.

[26] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli. Image denoising using scale mixtures of gaussians in the wavelet domain. *Image Processing, IEEE Transactions on*, 12(11):1338–1351, 2003.

[27] B. Recht, M. Fazel, and P. A. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM review*, 52(3):471–501, 2010.

[28] J. Sethuraman. A constructive definition of dirichlet priors. Technical report, DTIC Document, 1991.

[29] E. P. Simoncelli and E. H. Adelson. Noise removal via bayesian wavelet coring. In *Image Processing, 1996. Proceedings., International Conference on*, volume 1, pages 379–382. IEEE, 1996.

[30] J.-L. Starck, E. J. Candès, and D. L. Donoho. The curvelet transform for image denoising. *Image Processing, IEEE Transactions on*, 11(6):670–684, 2002.

[31] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *Computer Vision, 1998. Sixth International Conference on*, pages 839–846. IEEE, 1998.

[32] Y. Tsin, V. Ramesh, and T. Kanade. Statistical calibration of ccd imaging process. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, volume 1, pages 480–487. IEEE, 2001.

[33] R. Vidal. A tutorial on subspace clustering. *IEEE Signal Processing Magazine*, 28(2):52–68, 2010.

[34] Y.-Q. Wang and J.-M. Morel. Sure guided gaussian mixture image denoising. *SIAM Journal on Imaging Sciences*, 6(2):999–1034, 2013.

[35] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *Image Processing, IEEE Transactions on*, 13(4):600–612, 2004.

[36] G.-Z. Yang, P. Burger, D. N. Firmin, and S. Underwood. Structure adaptive anisotropic image filtering. *Image and Vision Computing*, 14(2):135–145, 1996.

[37] Q. Zhao, D. Meng, Z. Xu, W. Zuo, and L. Zhang. Robust principal component analysis with complex noise. In *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, pages 55–63, 2014.

[38] D. Zoran and Y. Weiss. From learning models of natural image patches to whole image restoration. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 479–486. IEEE, 2011.